

# On The Inclusion of Channel's Time Dependence in a Hidden Markov Model for Blind Channel Estimation

Carles Antón-Haro, José A. R. Fonollosa and Javier R. Fonollosa.  
Dpt. of Signal Theory and Communications. Universitat Politècnica de Catalunya.  
c/ Gran Capità s/n. 08034 Barcelona (SPAIN)  
Tel: +34-3-4016454, Fax: +34-3-4016447, e-mail: carles@gps.tsc.upc.es

## Abstract

In this paper we introduce several modifications to the Baum&Welch (BW) formulas used to reestimate the parameters of a Hidden Markov Model (HMM). The estimated parameter is the channel impulse response (CIR) of a communication system which is known to be time-varying. With these modifications, channel tracking properties of a BW-based algorithm are improved. The resulting algorithm is tested in a specific mobile radio environment (the GSM system), exhibiting good performance at expenses of higher computational complexity.

## 1. Introduction

It's well known that no high-speed band-limited digital communication can be carried out without the help of an equalizer. Conventional approaches to the adjustment of this equalizer require the transmission of a *training sequence* (i.e. known *a priori* by the receiver and the transmitter), which provides an accurate initial estimate for the equalizer taps; afterwards, slighter adjustments can be made on-line to adapt this first estimate to the, almost always, changing environment. Of course, the transmission of these training sequences, when possible, brings down the capacity of the system. For that reason, there is an increasing interest around blind equalizers [1,2,3] which deal with the problem of the adjustment *without* training sequences (i.e. blindly).

In [3], an Estimation-Modification (EM) Viterbi-based algorithm is proposed to perform jointly a Maximum Likelihood (ML) channel estimation and sequence detection. However, modelling the received signal as a HMM allows us to make use of the complete theory developed for these models. For example, the Baum&Welch (BW) algorithm was proposed in [7] to estimate the parameters of the channel and the characteristics of the modulation. This algorithm is known to lead, at least, to a local maximum of the likelihood function [4], what is not guaranteed by the Viterbi algorithm (VA). In this paper, several modifications

to this previously proposed algorithm are introduced to cope with the special features of mobile radio channels.

## 2. Signal model

As mentioned before, the environment in which the new algorithm is tested is the Pan-European Mobile Radio System, also known as GSM. In this system, a constant-envelope Gaussian Minimum Shift Keying (GMSK) modulation scheme with equivalent bandwidth (BT) equal to 0.3 is used. The access strategy is TDMA with 8 timeslots per carrier and 156.25 bit-intervals per timeslot in Normal bursts. At the chosen bit rate (270.8 kb/s), multipath propagation leads to deep fades and to uncontrolled Intersymbol Interference (ISI). Besides, and due to the mobile nature of the receiver, Doppler effect is also observed.

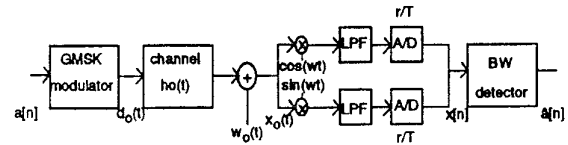


Fig. 1: Transmission subsystem.

Taking into account the above mentioned features, the signal at the input of the BW detector can be modelled as:

$$x[n] = f(s[n]) + w[n] \quad (1)$$

where  $f(\cdot)$  is a non-linear function of the present state  $s[n]$ , and  $\{w[n]\}$  denotes a sequence of zero-mean Gaussian variables with variance  $\sigma^2$  (AWGN). If we go on developing an expression for  $f(\cdot)$  we get:

$$f(s[n]) = \sum_{i=0}^{l_c-1} h[i] d[n-i] = \sum_{i=0}^{l_c-1} h[i] e^{j\phi[n-i]} \quad (2)$$

where  $\mathbf{h}$  and  $\mathbf{d}$  are the baseband equivalences for  $\mathbf{h}_0$  and  $\mathbf{d}_0$ . For a modulation index of 0.5,  $\phi[n]$  can be expressed as:

$$e_i^2 = \frac{\sum_{n=1}^L \gamma_i[n] |x[n] - m_i^{(0)} - m_i^{(1)} \cdot n|^2}{\sum_{n=1}^L \gamma_i[n]}, \quad 1 \leq i \leq N \quad (13)$$

Differentiating each component in the sum  $\varepsilon^2 = \sum e_i^2$  with respect to  $m_i^{(0)}$  and  $m_i^{(1)}$  we obtain:

$$\nabla e_i^2 = \begin{pmatrix} \frac{\partial e_i^2}{\partial m_i^{(0)}} \\ \frac{\partial e_i^2}{\partial m_i^{(1)}} \end{pmatrix} = -2 \cdot \begin{pmatrix} \Re \left\{ \sum_{n=1}^L \gamma_i[n] (x[n] - m_i^{(0)} - m_i^{(1)} \cdot n) \right\} \\ \Re \left\{ \sum_{n=1}^L \gamma_i[n] (x[n] - m_i^{(0)} - m_i^{(1)} \cdot n) \cdot n \right\} \end{pmatrix} \quad (14)$$

whose Hessian is positive definite unless

$$c1) \quad \sum_{n=1}^L \gamma_i[n] = 0 \quad (15)$$

(i.e. that state was not observed along the timeslot), or

$$c2) \quad \sum_{n=1}^L \gamma_i[n] = \gamma_i[n_i] \quad (16)$$

(i.e. that state was observed only once, when  $n=n_i$ ). In those cases, of course, there is no sense in looking for a linear approximation. From equating the gradient to zero and carrying out proper transformations, we find that

$$m_i^{(0)} = \frac{A \cdot \left( \sum_{n=1}^L \gamma_i[n] x[n] \cdot n \right) - B \cdot \left( \sum_{n=1}^L \gamma_i[n] x[n] \right)}{\Delta} \quad (17)$$

$$m_i^{(1)} = \frac{C \cdot \left( \sum_{n=1}^L \gamma_i[n] x[n] \right) - B \cdot \left( \sum_{n=1}^L \gamma_i[n] x[n] \cdot n \right)}{\Delta} \quad (18)$$

$$\hat{\sigma}^2 = \frac{1}{L} \sum_{n=1}^L \sum_{i=1}^N \gamma_i[n] |x[n] - m_i^{(0)} - m_i^{(1)} \cdot n|^2 \quad (19)$$

where

$$A = \sum_{n=1}^L \gamma_i[n]; \quad B = \sum_{n=1}^L \gamma_i[n] \cdot n; \quad C = \sum_{n=1}^L \gamma_i[n] \cdot n^2 \quad (20)$$

$$\Delta = A \cdot C - B^2$$

provide the components of the desired vector and an estimate for the variance of the AWGN.

Finally, special measures should be taken for the cases above mentioned in which  $m_i^{(0)}$  and  $m_i^{(1)}$  remain undefined. In the first case, *c1*), those components of the means vector are not considered in order to obtain  $\mathbf{h}$ . The method adopted is blocking them with a (diagonal) Weighting Matrix to be included in the LS estimate of step 3. The elements of such matrix are a measure of the reliability in the estimation of every component of vector  $\mathbf{m}$ , as a function of the times this state was observed along the sequence. To be precise:

$$\mathbf{W} = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & w_N \end{pmatrix} \quad w_i = \sum_{n=1}^L \gamma_i[n] \quad (21)$$

In the second case, *c2*), the *static* estimate for  $\mathbf{m}$  replaces the linear approximation. That is:

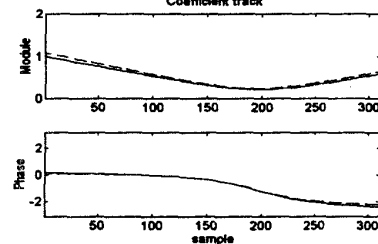
$$m_i^{(0)} = \frac{\sum_{n=1}^L \gamma_i[n] x[n]}{\sum_{n=1}^L \gamma_i[n]} = x[n_i] \quad (22)$$

$$m_i^{(1)} = 0$$

The resulting algorithm will be referred from now on as the Time-Dependent BW (TDBW) algorithm.

## 5. Simulation results

We tested the performance of the algorithm for the channels described in the ETSI recommendations. The speed for the mobiles in each environment was chosen according to [8]. Among all the cases, the most interesting ones were RA250 and RA100 (Rural Area Environment; speed equal to 100 and 250 km/h), since channel coherence time-intervals are the lowest ones. It should also be remarked that a sampling rate of 2 samples/symbol was considered to compensate for possible timing errors.

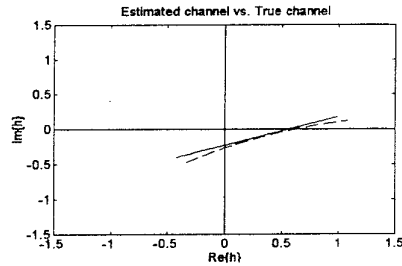


**Fig. 2:** Tracking for the first tap of a RA250 channel vs. time in amplitude and phase for the proposed algorithm. Dashed lines stand for the true channel; solid lines for the TDBW estimate.

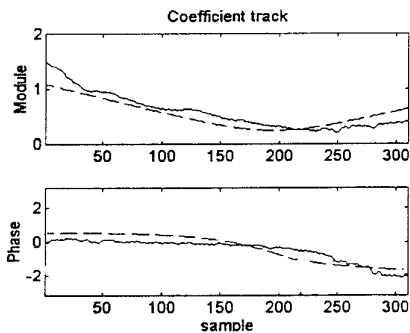
Channel tracking properties of the proposed algorithm are shown in Fig. 2 and Fig. 3. It can be observed that such properties are good as long as the linear approximation for the channel evolution is feasible. Comparing those figures with those obtained with the ABW algorithm (Fig. 4), we conclude that CIR tracking is now much less noisy.

In addition to this, now there is no need for waiting the algorithm to converge within the first samples of each timeslot. Moreover, the TDBW version is far more robust against deep fades which usually make the ABW algorithm to lose tracking. The reason for this robustness is that TDBW is a batch-type algorithm, where every sample in the

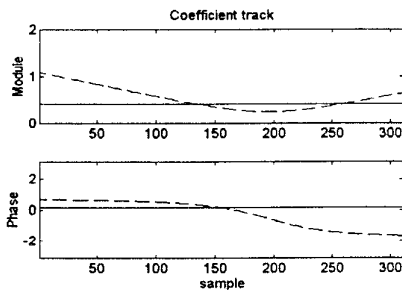
timeslot is used to estimate the CIR in every instant (even though in deep local fades), whereas in the ABW version the estimate relies mainly on the previous and, maybe, already-faded samples. Of course, those improvements are conditioned to an approximately linear variation of the channel, what is not required in the ABW algorithm. However, if this requirement was not met, it would always be possible to increase polynomial order to obtain a better approximation for the channel evolution.



**Fig. 3:** Tracking for the first tap of a RA250 channel in rectangular coordinates ( $\text{Re}\{h(t)\}$  and  $\text{Im}\{h(t)\}$ ). Dashed lines stand for the true channel; solid lines for the TDBW estimate.



**Fig. 4:** Tracking for the first tap of a RA250 channel vs. time. Dashed lines stand for the true channel; solid lines for the ABW estimate.



**Fig. 5:** Tracking for the first tap of a RA250 channel vs. time in amplitude and phase. Dashed lines stand for the true channel; solid lines for the BBW estimate.

The main advantage with respect to the BBW algorithm described in section 3, is the ability to track the evolution of the CIR along the timeslot, instead of approximating each

tap by a constant value (Fig. 5). For mobile stations exhibiting rather high speed, it reverts in a lower BER.

On the other hand, the main drawback of the proposed TDBW algorithm is the increase in the computational burden when compared with both the BBW and ABW versions. And, what is more, now the number of parameters to be estimated is double the quantity required before ( $\mathbf{m}^{(0)}$  and  $\mathbf{m}^{(1)}$  vs.  $\mathbf{m}$ ), whereas the amount of data available to perform that estimation is just the same (one timeslot). Consequently, for CIRs exhibiting large delay-spreads such as HT (Hilly Terrain environment), the variance increase in the estimation of some components in  $\mathbf{m}[n]$  is very severe and the Weighting Matrix cannot prevent the system from unstability. In those cases, the only way to make the algorithm to converge is considering larger timeslots which contain more symbols.

## 6. Conclusions

A new technique to include the time-varying nature of the parameters of a HMM in the BW reestimation formulas has been presented. The resulting algorithm for blind channel estimation, TDBW, has been compared with those proposed in previous references (BBW and ABW), and its performance qualitatively evaluated in a very concrete environment (the GSM system). The most important drawback of the algorithm is its high computational cost.

Future work is concerned about applying the theory of HMMs and the developed BW-based algorithms to other communication environments such as Underwater Acoustics (UWA), or in other communication problems.

## References

- [1] J.G. Proakis and C.L. Nikias, "Blind Equalization", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 76-87, 1991.
- [2] John Shynk et al., "A Comparative Performance Study of Several Blind Equalization Algorithms", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 102-117, 1991.
- [3] M. Ghosh and C.L. Weber, "Maximum Likelihood Blind Equalization", *Proc. SPIE Adaptive Signal Processing* 1991, Vol. 1565, pp. 188-195, 1991.
- [4] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. IEEE*, Vol. 77, No. 2, pp. 257-286, Feb. 1989.
- [5] R. Steele, "Mobile Radio Communications", *Pentech Press Publishers*, London 1992.
- [6] T. Aulin and C.W. Sundberg, "Continuous Phase Modulations (Parts I and II)", *IEEE Trans. on Comm.*, Vol COM-29, No. 3, pp. 196-225, March 1981.
- [7] José A.R. Fonollosa and J. Vidal, "Application of Hidden Markov Models to Blind Channel Characterization and Data Detection", *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, pp. 185-188, Australia, April 1994.

$$\phi[n] = \pi \sum_{r=-R}^R q[r] a[n-r] + \theta[n] \quad (3)$$

where  $q[r] \in [0, 0.5]$  are the weights corresponding to the (sampled) gaussian-shaping pulse and  $\theta[n] \in \{0, \pi/2, \pi, 3\pi/2\}$  accounts for the accumulated phase at instant  $n$  [5,6]. Now we conclude that the number of transmitter symbols (bits) involved in a single observation at the receiver is given by:

$$l_t = l_m + l_c - 1 = (2R+1) + l_c - 1 \quad (4)$$

However, the amount of ISI produced by the GMSK modulator for BT=0.3 can be neglected without significant performance loss. Under this simplifying assumption, which reverts in a lower number of states and a reduced computational complexity, we get that:

$$R=0 \Rightarrow l_t = l_c \quad (5)$$

At this point, we can already model each observation in the received sequence,  $\mathbf{x}_L = (x[1], x[2], \dots, x[L])^T$ , as a probabilistic function of the present state  $\mathbf{s}[n] = (a[n], \dots, a[n-l_t+1], \theta[n])^T$ , obtaining a description of  $\mathbf{x}_L$  as a first order HMM with  $N=4 \cdot 2^H$  states.

### 3. Overview of the BW algorithm

On the basis of this first order HMM and by means of the BW algorithm, it is possible to obtain a solution to the problem of the identification of the unknown parameters of the model, that is,  $\sigma^2$  and  $\mathbf{h} = (h_0, \dots, h_{l_c-1})^T$  [7]. To be precise, the parameters really estimated are  $\hat{\sigma}^2$  and the means vector  $\mathbf{m} = (m_1, \dots, m_N)^T$  corresponding to the noise-free ISI-corrupted received signal associated with the  $N$  states of the system. Assuming a FIR model for the channel,  $\mathbf{m}$  is related to  $\mathbf{h}$  through the linear constraints:

$$\begin{aligned} \hat{\mathbf{m}} &= \mathbf{D} \hat{\mathbf{h}} \\ \hat{\mathbf{h}} &= \mathbf{D}^\# \hat{\mathbf{m}} \end{aligned} \quad (6)$$

where  $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N)^T$  is a  $N \times l_c$  full-rank matrix containing in its rows all the  $l_c$ -tuples,  $\mathbf{d}_i = (d_i^{(1)}, d_i^{(2)}, \dots, d_i^{(l_c)})^T$ , corresponding to the modulator consecutive outputs associated to the  $N$  different states of the system.  $\mathbf{D}^\#$  denotes pseudoinverse. The batch Baum&Welch (BBW) algorithm, thoroughly explained in [7], can be outlined as:

1. Projection of  $\mathbf{h}$  on  $\mathbf{m}$  by means of the additional linear constraints:

$$\hat{\mathbf{m}} = \mathbf{D} \hat{\mathbf{h}} \quad (7)$$

2. Reestimation of  $\sigma^2$  and  $\mathbf{m}$  using the BW reestimation formulas.

3. Least Squares (LS) estimation of  $\mathbf{h}$  using again linear constraints.

$$\hat{\mathbf{h}} = \mathbf{D}^\# \hat{\mathbf{m}} \quad (8)$$

4. Repeat steps 1..4 until convergence.

The BW reestimation formulas used in step 2, state as follows:

$$\begin{aligned} \hat{m}_i &= \frac{\sum_{n=1}^L \gamma_i[n] x[n]}{\sum_{n=1}^L \gamma_i[n]}, \quad 1 \leq i \leq N \\ \hat{\sigma}^2 &= \frac{1}{L} \sum_{n=1}^L \sum_{i=1}^N \gamma_i[n] |x[n] - \hat{m}_i|^2 \end{aligned} \quad (9)$$

where  $\gamma_i[n]$  is the probability of being in state  $i$  at instant  $n$  given the model and the observed sequence. However, these formulas implicitly assume the CIR to be stationary within a timeslot duration, what is not realistic when the timeslot is long enough or the channel varies rapidly. Hence, we will obtain other reestimation formulas to solve this problem.

### 4. Modified algorithm

Several strategies can be considered to cope with the time-varying nature of the channel: estimating the CIR with a recursive adaptation scheme such as LMS what is referred in [7] as the Adaptive BW (ABW) algorithm; or fragmenting each timeslot in subblocks. As opposed to those methods, we will try to include the time-varying nature of the channel directly in the reestimation formulas.

We can approximate the evolution of every tap in the CIR,  $h_j$ , by means of a polynomial in  $n$ :

$$\hat{h}_j[n] = h_j^{(0)} + h_j^{(1)} \cdot n + h_j^{(2)} \cdot n^2 + \dots \quad (10)$$

For the channels specified in the ETSI recommendations and assuming the speed of the mobile to be less than 250 km/h, the linear approximation was observed to be good enough. Assembling all the taps in a single column vector:

$$\hat{\mathbf{h}}[n] = \mathbf{h}^{(0)} + \mathbf{h}^{(1)} \cdot n \quad (11)$$

and applying the linear transformation above described

$$\hat{\mathbf{m}}[n] = \mathbf{D} \hat{\mathbf{h}}[n] = \mathbf{D} (\mathbf{h}^{(0)} + \mathbf{h}^{(1)} \cdot n) = \mathbf{m}^{(0)} + \mathbf{m}^{(1)} \cdot n \quad (12)$$

we observe that the evolution for the means is also linear. Vectors  $\mathbf{m}^{(0)}$  and  $\mathbf{m}^{(1)}$  will be obtained in order to minimize the MSE given by the following expression: